



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Teaching a Text Mining Bootcamp in Lockdown

Citation for published version:

Llewellyn, C, Orzechowski, P & Alex, B 2020 'Teaching a Text Mining Bootcamp in Lockdown' Edinburgh, pp. 1-7.

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Teaching a Text Mining Bootcamp in Lockdown

Clare Llewellyn, Pawel Orzechowski and Beatrice Alex

June 2020



It feels like we are all in crisis mode at the moment. We want to teach but because of social distancing measures we can't do this in a physical classroom. We need to learn new ways of interacting and a multitude of different technologies and we need to do it fast. We had been thinking about this for a while before Covid-19 hit. We were designing a technologically innovative course but clearly because of the situation we had to adapt the way we delivered it. Rather than hybrid teaching as intended we had to fully commit to online methods. We decided it was now or never and we should plunge headlong into the digital teaching world.

The easiest way to do this would have been to post videos of a traditional style lecture – it is very tempting to take this approach. But, it is important that we maintain what is good about teaching when you are all in the same room, the collaboration, its social aspects, the feedback all of which you lose if a student sits on their own in a room watching a pre-recorded lecture.

We decided to run a Text and Data Mining boot camp to virtually test a new course we are planning as part of the Edinburgh Futures Institute postgraduate Futures Programmes (it was run on this occasion in conjunction with the Business School's Student Development team). We wanted to not only teach students but also to teach ourselves how to become better practitioners in an online world.

A Virtual Learning Experience

In our data-driven society, it is increasingly essential for people throughout the private, public and third sectors to know how to analyse the wealth of information society creates each day. The Text and Data Mining Boot Camp gives those who have very limited or no coding experience the tools they need to interrogate data. This course is designed to teach non-coders how to analyse textual data using the Python programming language. It takes those who have no or very limited coding experience through all the required steps needed to build an analysis and impressive visualisations of large sets of documents. We ran this over three afternoon sessions on a Monday, Wednesday and a Friday, with an office hour on Tuesday and Thursday to sort out any issues. It was hoped that by the end of the course the participants would have text mining skills they could use in their own research.

We wanted to test the content of this course but also the methods for teaching. We are all likely to be teaching virtually more often in the future. For example, the Edinburgh Futures Institute plans to run hybrid courses to students across the world. Pawel is a strong proponent of pair programming, where two students work together on a single machine to solve problems. This allows the students to learn from each other as well as from their teacher. We wanted to see if it was possible to take this approach into a virtual teaching environment. In addition to the students learning the skills it also provides a social interaction which is much needed at the moment!

Each session introduced participants to a new topic such as Python skills for reading and processing text and visualisations of large datasets, through a live video lecture. The topics built upon the knowledge from the previous sessions. In the first afternoon we started with Python skills for reading in and processing text. On the second day we looked at using much larger sets of text and added in visualisations. We used two data sets, the Medical History of British India provided by the National Library of Scotland¹ and the inaugural addresses of all American Presidents from 1789 to 2017 (see Figure 1). During the third session we looked at part of speech tagging and labelling and extracting specific entities such as person names from the text.

The backgrounds of those who signed up for the course were completely mixed. Everyone was either a student or a member of staff at Edinburgh University, but there was every level from professors to undergraduates, from around the world. Some were in different time zones and participants came from many different schools within the University.

¹<https://data.nls.uk>

and raising questions to the speaker. This was something we found was essential as it was very easy to lose flow and get distracted without this help. We found that it isn't really possible to teach live online on your own.

Each day the students were given two sets of worked through problems using Noteable. Noteable is a platform that creates a virtual coding environment for each student, it is based on the Jupyter Notebook technology. This is a really important resource as it reduces the need for the student to download and set up software. They use Noteable directly through Learn⁶ virtual learning environment (VLE) in their own browser – it works best in Chrome. The students were given a link to a Github repository from which they could pull new notebooks at the beginning of each session. The notebooks were a mix of explanation, code to run and mini or extended programming tasks. For each hour-long session students were assigned a random buddy and they were put in a group within a Collaborate video call. This feels a bit like being put in a separate room with your buddy. You can chat and share screens without being overheard by other people. If the students got stuck, they could raise a virtual hand and an instructor could drop into the room to help and answer questions or resolve programming issues. We also randomly popped into the rooms to see how everyone was doing.

After the first session we pulled everyone back into the shared room and asked the students to fill in a quick survey to give us feedback. We answered any questions, had a quick break, then it was on to the second notebook with a new buddy. We wrapped up the day with a quick chat and another session of feedback.

Relentless Feedback

This relentless feedback loop helped us realise we could cover more content as time went on, it is one of the reasons we believe we had such a high participant retention rate.

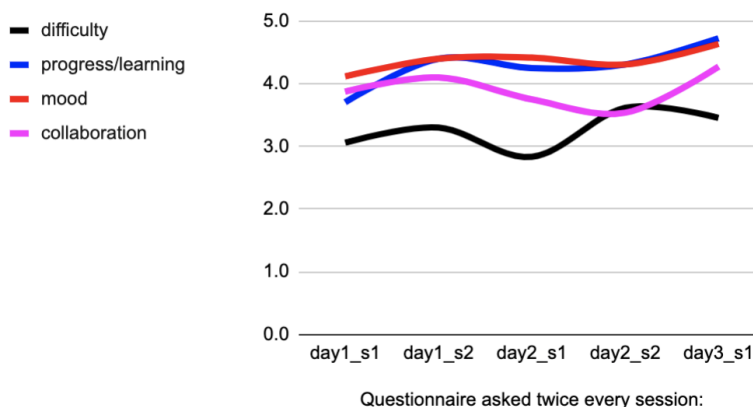
We decided to change the format for the second day based on the feedback we had been given. The students requested a quick recap of the first day, which we hadn't planned on doing but it was a great way to link the sessions and get everything fresh in everyone's minds. We also worked through the first section of that day's notebook in class, so everyone had a clear idea of what to do. We were very pleased that we had a very high retention rate for students. We had lost a couple of participants in the first hour of the first session but everyone who finished the first day came back to the second.

We found it was really important to be flexible and adapt to what the students wanted. The twice-a-day mini-feedback form was really helpful for that made it very clear which parts of the course on day 2 and 3 were in response to their feedback (see feedback analysis in Figure 2). Still, we all found teaching in an online environment quite odd. You lose the sense of whether the students are learning and enjoying the experience because you can't see their faces or body language. The feedback does help, even in something as simple as 'raise your hand if you can hear me', but it is still odd to talk to a screen without seeing everyone.

⁶<https://www.learn.ed.ac.uk>

	AVG	MODE	day1_s1	day1_s2	day2_s1	day2_s2	day3_s1
difficulty	3.2	3	3.1	3.3	2.8	3.6	3.5
progress/learning	4.2	4	3.7	4.4	4.3	4.3	4.7
mood	4.3	5	4.1	4.4	4.4	4.3	4.6
collaboration	3.9	5	3.9	4.1	3.8	3.5	4.3

Learning KPIs across 3 days



“Fantastic! The pair learning is excellent! Jupiter notebooks are a great tool. The real time interactivity is super rewarding.”

Figure 2: Feedback analysis for all surveys over the course of the boot camp and a quote from one student (with permission to share).

We also found that the amount of content that we could cover grew as the time went on. There were issues with the technology at first which needed fixing and we were all getting used to the new technology. Hopefully this lag would be lower when students and teachers have completed a course. The conversation became more natural as time went on. At first it was quite odd to drop into the rooms but by the second day we were all chatting a lot more. The students really liked the pair programming, they liked the flexibility and the content. They really felt they were part of the course in a way you don’t always get online.

A Measurable Impact

We didn’t get everything right. The technology didn’t always work and luckily Pawel is quite experienced in fixing these types of issues particular to the software. If Pawel

hadn't had this experience it would have been much harder. We were slowed down on the first day as some of the students hadn't done the pre-course work. We should have made that compulsory or taught it in the class. We didn't give enough thoughts to accessibility; we just assumed the software would deal with that - it didn't. We should have asked all students beforehand and made time to deal with any issues that arose. The students were a bit shy when it came to talking to each other and putting on their webcams. In the future we would have more ice breaking upfront.

There are some quite simple things that made a lot of difference. We played music in the room before the class so when students joined, they knew we were there and that their speakers were on. We made use of the virtual white board a lot to gather really fast anonymous feedback in addition to frequent short surveys (see Figure 3). We had questions in the notebooks that the buddies had to work on together to encourage discussion. We had essential tasks and tasks for the curious that were more complex. This allowed some students to extend their learning without others feeling they were left behind.

We had great feedback on the course. On participant thought it was "Fantastic!" in their final feedback survey. Another thought "The pair learning is excellent! Jupiter [sic] notebooks are a great tool. The real-time interactivity is super rewarding." Others reported that the three lecturers and the "humour and playfulness of the examples" made the course "really great, especially for someone completely new to coding." While another commented they would use the skills they learned in gathering data for their undergraduate dissertation.

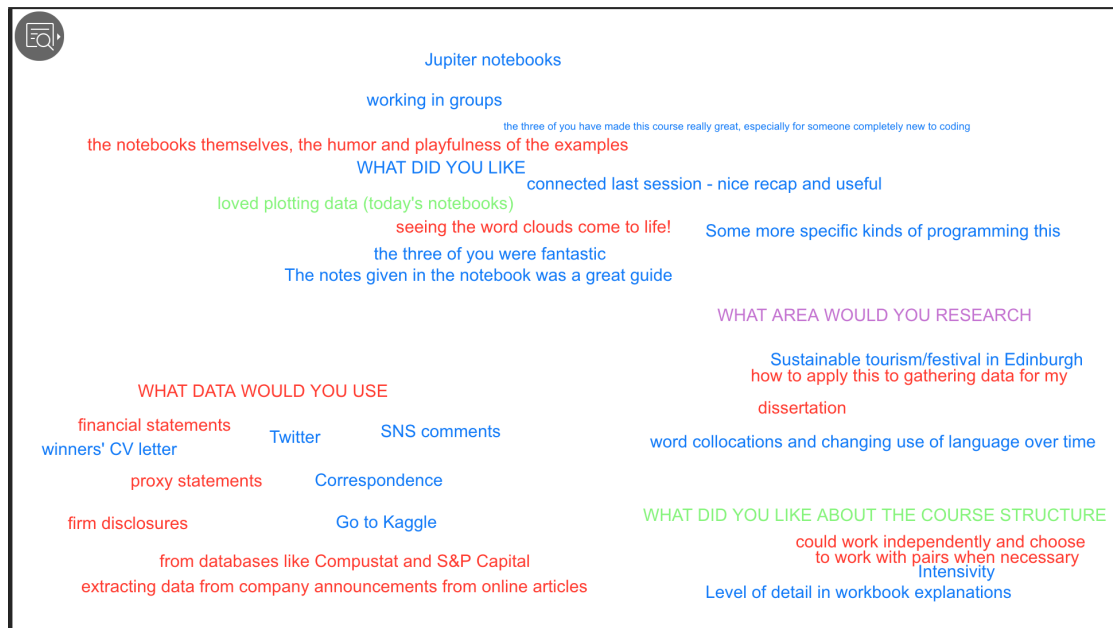


Figure 3: White board with feedback generated by students in the course.

As instructors, we found that teaching in this way, switching between modes, lecturing,

answering the chat, live coding, responding to issues is really cognitively challenging. It is hard work and can't be done by one individual. The technology is complex and can fail but it is really intuitive and provides a wide range of ways to teach and interact. We learned that online teaching is exhausting but if you do it right it can still be really rewarding. We all enjoyed the interactions and felt part of a little community. After the course we did a debriefing and each wrote down three things we liked about the course and something we wished we could have achieved (see page 8).

For the Future

Going forward we'd like to experiment with teaching this course in different ways: asynchronously to students joining from different time zones, to much larger groups to understand where the limits are with number of participants given staff capacity, or in a writer-retreat type setup where the instructors touch base with students several times throughout the day. We will also look at how this course can be pivoted back to on-campus teaching for students who can join in person and in the event that the current pandemic is over. The lockdown encouraged us to innovate, and our experience demonstrates what is possible to achieve despite the limitations. Experiencing the learning in a classroom or lecture theatre is difficult to replicate online, however, we are confident these types of virtual environments will play a role in education beyond this pandemic, to complement and enhance traditional learning.

Acknowledgements

We would like to thank Lauren Porter and Siobhan Dunn at the Edinburgh Futures Institute for managing the registration for this course and Marco Rossi at the Business School's Student Development Team for offering the course to their students. We would also like to thank Prof. Laura Cram for supporting us in the long-term development of the EFI course as part of EFI's PGT training. This bootcamp was built on and expands material developed for a Library Carpentries course with the support of the Centre for Data, Culture and Society.

Biographies

Dr Clare Llewellyn is Career Development Fellow at The University of Edinburgh, School of Social and Political Science.

Dr Pawel Orzechowski is Senior Teaching Fellow in Programming for Business at the Business School.

Dr Beatrice Alex is a Chancellor's Fellow and Turing Fellow, and works across EFI, the School of Literatures, Languages and Cultures and the School of Informatics.

Three stars and a wish

★ I liked that we were all willing to try anything and go beyond our comfort zone and fail

★ I liked the way we naturally supported each other and took different roles, and swapped those roles

★ I enjoyed learning the technology – something I thought I'd hate!

I wish we could find a way to make the students more interactive, chat, turn their cameras on.



★ Relentless feedback: asking 30s pulse-checking questionnaires twice a day; we used Office Forms.

★ Staff room: easy and persisting internal comms channel within the course development team; we used Teams.

★ Poetic licence: the person who created the final version of the notes was allowed to make adjustments to the content, so they fit the format.

Our greatest ally was the tech that just worked. I wish we could also provide core hybrid experience of being in the same room.

★ Great team of people, great combination of skills

★ Brilliant to see feedback from students coming in

★ I learned a lot about teaching online

I wish I could witness the learning better online.

